ABSTRAK

ANALISIS KOMPONEN UTAMA

Mariana, Dosen Pendidikan Matematika
Fakultas Tarbiyah dan Keguruan, IAIN Ambon
085244357173, E-mail: anastt_10@yahoo.com

Mengatasi masalah multikolinieritas pada analisis regresi linier berganda dapat menggunakan analisis komponen utama (*principal component analysis*). Penggunaan analisis komponen utama bertujuan untuk mereduksi dimensi data yang saling berkorelasi menjadi dimensi data yang tidak saling berkorelasi yaitu variabel-variabel baru yang saling bebas atau tidak berkorelasi. Regresi komponen utama cukup efektif dalam mengatasi masalah multikolinearitas. Ini terlihat dimana nilai VIF pada regresi komponen utama bernilai satu hal ini menunjukkan tidak terdapat korelasi antar variabel komponen utama.

Kata Kunci : Komponen Utama

PENDAHULUAN

Analisis regresi linear merupakan suatu teknik statistika yang digunakan untuk menjelaskan pengaruh variabel bebas (independent variable) terhadap variabel tak bebas (dependent variable). Analisis regresi yang hanya melibatkan satu variable bebas di sebut Analisis regresi linear sederhana, sedangkan yang melibatkan lebih dari satu variabel bebas disebut analisi regresi linear berganda. Salah satu asumsi yang harus dipenuhi untuk melakukan pengujian hipotesis terhadap parameter pada analisis regresi linier berganda adalah tidak terjadinya korelasi antar variabel bebas (multikolinier). Jika antar variabel saling berkorelasi tinggi, penggunaan Metode Kuadrat Terkecil (MKT) untuk menduga koefisien regresi menjadi tidak valid karena tidak dipenuhinya salah satu asumsi (Aunuddin, 1989). Akibatnya, hipotesis menunjukkan variabel-variabel bebas yang seharusnya berpengaruh signifikan terhadap variabel tak bebas akan dinyatakan sebaliknya (tidak nyata secara statistik), tanda koefisien regresi dugaan yang dihasilkan bertentangan dengan kondisi aktual, penduga koefisien regresi bersifat tidak stabil sehingga mengakibatkan sulitnya menduga nilai-nilai variabel tak bebas yang tentunya akan mengakibatkan tidak akuratnya pada pendugaan

(Myers, 1991). Indikasi multikolinieritas, salah satunya, dapat dideteksi dari *Variance Inflation Factor* (VIF).

Kondisi ini mendorong untuk dikembangkannya suatu cara atau teknik yang dapat digunakan untuk mengatasi masalah multikolinieritas pada analisis regresi linier berganda. Salah satu solusi yang dapat digunakan adalah dengan menggunakan analisis komponen utama (principal component analysis). Penggunaan analisis komponen utama bertujuan untuk mereduksi dimensi data yang saling berkorelasi menjadi dimensi data yang tidak saling berkorelasi yaitu variabel-variabel baru yang saling bebas (tidak berkorelasi). Variable-variabel baru tersebut merupakan kombinasi linier dari variabel-variabel bebas asal. Variabel-variabel baru yang dihasilkan itulah yang kemudian disebut komponen utama, dan selanjutnya diregresikan dengan variabel tak bebas. Penulisan makalah ini bertujuan untuk mengkaji analisis regresi komponen utama sebagai salah satu solusi dalam menangani multikolinieritas antar variabel bebas pada analisis regresi linier berganda, selanjutnya akan diberikan ilustrasi penerapan analisis regresi komponen utama dalam contoh kasus.

HASIL DAN PEMBAHASAN

1. Analisis Regresi Linier Berganda

Analisis regresi linier berganda merupakan analisis regresi yang melibatkan lebih dari satu variable bebas ($X_1, X_2, \dots , X_n$) yang disebut prediktor dan mempunyai hubungan linier dengan variabel tak bebas (Y) yang disebut respon. Model regresi linier berganda yang melibatkan *p* variabel bebas secara umum dinyatakan sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \varepsilon_i \text{ ; } i = 1, 2, \dots , n$$

Dimana Y adalah variabel tak bebas, X_i adalah variabel bebas ke-*i*, ε_i adalah sisaan dan $\beta_0, \beta_1, \beta_2, \dots , \beta_n$ adalah koefisien regresi. Model persamaan regresi secara umum juga dapat ditulis dalam notasi matriks sebagai berikut :

$$\mathbf{Y = X\beta + \varepsilon}$$

dimana Y adalah vektor peubah respon berukuran $n \times 1$ yang elemennya merupakan nilai-nilai amatan dari variabel tak bebas, X adalah matriks peubah bebas yang berukuran $n \times (p+1)$, $\boldsymbol{\beta}$ adalah vektor koefisien regresi yang berukuran $(p+1) \times 1$ dan $\boldsymbol{\varepsilon}$ adalah vektor galat berukuran $n \times 1$, dimana asumsi untuk ε_i yaitu :

1. $E(\varepsilon_i) = 0$ untuk i = 1, 2, … , n atau ini ekuivalen dengan
   $E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_n X_{ni}$
2. $Var(\varepsilon_i) = \sigma^2$ untuk i = 1, 2, … , n
3. $cov(\varepsilon_i, \varepsilon_j) = 0$ untuk setiap $i \neq j$

Parameter $\boldsymbol{\beta}$ biasanya diduga menggunakan metode kuadrat terkecil. Prinsip dasar metode ini yaitu meminimumkan jumlah kuadrat galat (JKG minimum). Menurut (Myers & Milton, 1991) Asumsi yang harus dipenuhi yaitu :

1. X adalah matriks non-singular (berpangkat penuh) atau tidak ada korelasi yang erat diantara peubah-peubah bebas $(cor(x_i, x_j) = 0, \forall i \neq j)$
2. $\boldsymbol{\varepsilon}$ adalah vektor acak dengan rataan 0 dan ragam σ^2 , ini bararti tidak ada autokorelasi antar galat $(\text{cov}(e_i, e_j) = 0, \forall i \neq j)$

Sehingga diperoleh penduga MKT sebagai berikut :

$$\hat{\beta} = (X'X)^{-1} X'Y$$

dimana $\hat{\beta}$ adalah penduga yang memenuhi sifat linear, tidak berbias dan memiliki ragam minimum.

Pada analisis regresi, salah saatu tujuan yang ingin dicapai adalah pengujian hipotesis terhadap koefisien regresi dengan tujuan untuk mengetahui kontribusi relative dari masing-masing variabel bebas. Pada MKT uji parameter regresi dapat dilakukan secara parsial menggunakan uji t. Bentuk umum uji hipotesisnya sebagai berikut :

$H_0 : \beta_j = 0$ artinya koefisien ke-j tidak signifikan atau variable bebas ke-j tidak berpengaruh nyata terhadap Y.

$H_0 : \beta_j \neq 0$ artinya koefisien ke-j signifikan atau variable bebas ke-j berpengaruh nyata terhadap Y.

Statistik uji yang digunakan untuk menguji parameter regresi secara parsial yaitu :

$$t_{hit}(\hat{\beta}_j) = \frac{\hat{\beta}_j}{\sqrt{\text{var}(\hat{\beta}_j)}}$$

Dengan kaidah keputusan jika $\left|t_{hit}(\hat{\beta}_j)\right| > t_{(n-p-1);\alpha/2}$, maka H_0 ditolak yang artinya variable bebas ke-*j* berpengaruh nyata terhadap *Y.*

2. Multikolinieritas

Multikolinieritas adalah suatu kondisi dimana terjadi korelasi antara variable-variabel bebas dalam regresi linear ganda. Akibatnya salah satu asumsi untuk menduga koefisien regresi menggunakan MKT tidak terpenuhi sehingga penggunaannya menjadi tidak valid (Aunnudin, 1989). Jika dilakukan untuk melakukan prediksi, model yang didapat akan menghasilkan prediksi yang buruk (menyimpang dari nilai aslinya). selain itu menurut jollife (1986) masalah multikolinearitas juga akan mengakibatkan :

- Koefisien regresi dugaan tidak nyata walaupun nilai R^2-nya tinggi.
- Nilai dugaan koefisien regresi sangat sensitive terhadap perubahan data.
- Dengan MKT, simpangan baku koefisien regresi dugaan sangat besar.

Salah satu metode formal yang dapat digunakan untuk mendeteksi adanya multikolinier adalah melalui faktor inflasi ragam (*Variance Inflation Factor*/VIF). VIF digunakan sebagai kriteria untuk mendeteksi multikolinieritas pada regresi linier yang melibatkan lebih dari dua variabel bebas. VIF_i memiliki persamaan sebagaiberikut :

$$VIF_i = \frac{1}{1-R_i^2}$$

dimana R_i^2 adalah koefisien determinasi dari regresi variabel bebas ke-*i* (X_i) dengan variabel bebas lainnya. Indikasi adanya masalah multikolinearitas yaitu jika nilai VIF lebih besar dari 10. Selain itu multikolinearitas dapat pula dideteksi dengan melihat akar ciri dari X'X. Jika ada satu atau lebih akar ciri bernilai kecil bahkan hampir nol berarti ada satu atau lebih hubungan linear yang erat antar peubah bebas.

3. Analisis Komponen Utama

Analisis komponen utama merupakan teknik statistik yang dapat digunakan untuk menjelaskan struktur ragam-peragam dari sekumpulan variabel melalui beberapa variabel baru dimana variabel baru ini saling bebas, dan merupakan kombinasi linier dari variabel asalnya. Variabel baru tersebut dinamakan komponen utama (*principal component*). Secara umum tujuan dari analisis komponen utama adalah mereduksi dimensi data yang besar dan saling berkorelasi menjadi dimensi data yang kecil dan tidak saling berkorelasi (jolliffe 2002), hal ini dilakukan untuk kebutuhan interpretasi.

Komponen utama dapat dibentuk dari matriks ragam-peragam ($\sum$) maupun matriks korelasi. Kedua matriks tersebut berguna dalam penghitungan nilai akar ciri λ_i dan vektor ciri γ_i. $\lambda_1, \lambda_2, ..., \lambda_p$ merupakan akar ciri yang diperoleh dari persamaan | $\sum$ - λI | = 0, sedangkan $\gamma_1, \gamma_2, ..., \gamma_p$ merupakan vektor ciri yang diperoleh dari persamaan $(\sum - \lambda_i I)\gamma_i = 0; i = 1,2,....p$. Misalkan X_1, X_2 , …….., X_p adalah peubah acak yang menyebar menurut sebaran tertentu dengan verktor nilai tengah μ serta memiliki pasangan akar ciri dan vektor ciri yang saling ortonormal $(\lambda_1, \gamma_1), (\lambda_2, \gamma_2), \ldots\ldots, (\lambda_p, \gamma_p)$, maka komponen utama ke-i dapat didefinisikan sebagai berikut :

$$KU_i = \gamma_{i1}' X_1 + \ldots\ldots\ldots + \gamma_{ip}' X_p$$

Berdasarkan definisi diatas ragam dari komponen utama ke-i adalah

$$\sigma^2_{KU1} = Var(KU_i) = \lambda_i = \gamma'_1 \sum \gamma_1 = \sum_{i=1}^{p}\sum_{j=1}^{p} \gamma_{1i}\gamma_{1j}\sigma_{ij}; j = 1,2,....., p$$

Hasil penurunan persamaan langrange menunjukkan bahwa λ_i merupakan akar ciri terbesar yang memaksimumkan ragam KU_1 dan γ_1 merupakan vektor ciri yang berpadanan dengan λ_i. KU_2 adalah komponen utama ke-2 yang memaksimumkan nilai $\gamma'_2 \sum \gamma_2 = \lambda_2$. KU_p adalah komponen utama ke-p yang memaksimumkan ragam KU_p dengan memaksimumkan $\gamma'_p \sum \gamma_p = \lambda_p$. Urutan KU_1, KU_2, …,KU_p harus memenuhi persyaratan $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$. Sementara itu, kontribusi keragaman dari setiap komponen utama ke-*k* terhadap total keragaman adalah

$$(\text{proporsi}) = \frac{\lambda_k}{tr(\Sigma)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + ... + \lambda p}$$

Matriks peragam digunakan bila semua peubah yang diamati diukur dalam satuan pengukuran yang sama, tetapi bila peubah yang diamati mempunyai satuan pengukuran yang berbeda, maka digunakan matriks korelasi, dalam hal ini veriabel bebas perlu dibakukan terlebih dahulu dalam variabel baku sebagai berikut :

$$z_p = \frac{x_p - \mu_p}{\sigma_p}$$

Korelasi antara peubah ke-*i* dengan komponen utama ke-*j* jika diturunkan berdasarkan matriks peragam dinyatakan sebagai $r_{x_i y_j} = \frac{\gamma_i \sqrt{\lambda_j}}{s_i}$ dengan λ_j adalah akar ciri matriks peragam dan s_i adalah simpangan baku peubah ke-*i*. Sedangkan jika diturunkan berdasarkan matriks korelasi maka $r_{x_i y_j} = \gamma_i \sqrt{\lambda_j}$.

4. Kriteria Pemilihan Komponen Utama

Salah satu tujuan dari analisis komponen utama adalah mereduksi dimensi data asal yang semula terdapat *p* variabel bebas menjadi *k* komponen utama (dimana $k < p$). Langkah awal yaitu menghitung skor masing-masing komponen utama. Lalu dipilih k komponen ($k < m$) untuk digunakan sebagai peubah bebas dalam MKT. Secara umum kriteria pemilihan *k* komponen utama yaitu :

1. Dalam pemilihan jumlah komponen teresebut belum ada aturan tertentu yang disepakati oleh semua ahli statistika. Sebagian ahli statistika ada yang mengambil akar ciri yang lebih besar dari 1 atau mengambil komponen utama tertentu, dimana proporsi keragaman y yang dapat diterangkan oleh komponen tersebut dianggap cukup berarti.

2. Proporsi kumulatif keragaman data asal yang dijelaskan oleh k komponen utama minimal 80%, dan proporsi total variansi populasi bernilai cukup besar.
3. Dengan menggunakan scree plot yaitu plot antara i dengan λ_i, pemilihan nilai k berdasarkan scree plot ditentukan dengan melihat letak terjadinya belokan dengan menghapus komponen utama yang menghasilkan beberapa nilai eigen kecil membentuk pola garis lurus.

5. Analisis Regresi Komponen Utama

Metode regresi komponen utama merupakan teknik analisis komponen utama yang dikombinasikan dengan teknik regresi MKT. Prinsipnya yaitu dengan memilih beberapa komponen utama pertama yang akan digunakan sebagai peubah bebas dalam regresi MKT. Dalam hal ini, jika semua komponen utama digunakan sebagai peubah bebas, maka akan dihasilkan model yang setara dengan yang diperoleh melalui MKT (Jollife, 1986).

Prosedurnya diawali dengan melakukan pembakuan peubah bebas. Misalkan matrik Z berasal dari matriks X yang terpusatkan dan terskalakan, yaitu :

$$z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{s_j^{1/2}}$$

Dimana $s_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2, i = 1,2,\ldots\ldots\ldots\ldots, n$ dan, $j = 1,2,\ldots\ldots\ldots\ldots, m$.

Maka matriks korelasinya adalah Z'Z dan akar ciri dari matriks korelasinya yaitu $\lambda_1 > \lambda_2 > \ldots\ldots. > \lambda_j$ diperoleh dari persamaan determinan :

$$| Z'Z - \lambda_j I | = 0$$

Untuk setiap akar ciri λ_j terdapat vektor ciri γ_j yang memenuhi :

$$(Z'Z - \lambda_j I) \gamma_j = 0$$

Vektor cirinya yaitu $\gamma_j = (\gamma_{1j}, \gamma_{2j}, \ldots\ldots.., \gamma_{mj})'$ merupakan solusi ternormalkan sedemikian rupa sehingga $\gamma_j' \gamma_j = 1$

Fungsi komponen utama KU_j merupakan kombinasi linear antara matriks Z dengan vektor γ_j dalam bentuk :

$$KU_j = \gamma_{1j}Z_1 + \gamma_{2j}Z_2 + \dots\dots + \gamma_{mj}Z_m\,\gamma \quad \dots\dots\dots\dots\dots\dots\dots(1)$$

Sehingga persamaan regresi komponen utama yang didapat adalah sebagai berikut :

$$Y = \delta_0 + \delta_1 KU_1 + \delta_2 KU_2 + \dots\dots + \delta_k KU_k \quad \dots\dots\dots\dots(2)$$

Dari k komponen utama, missal diambil g komponen utama, selanjutnya berdasarkan persamaan (1) dan (2), persamaan regresi komponen utama dapat ditransformasikan ke peubah asal yang dibakukan yaitu :

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \dots\dots + \beta_m Z_m \quad \dots\dots\dots\dots\dots\dots\dots(3)$$

Dimana $\beta_0 = \delta_0$

$$\beta_j = \gamma_{j1}\delta_1 + \gamma_{j2}\delta_2 + \dots\dots + \gamma_{jk}\delta_k$$

$$j = 1,2,3,\dots\dots m$$

Ragam koefisien regresi komponen utama dihitung dengan rumus :

$$Var(PC_g) = s^{*2}\sum_{g=1}^{m}\frac{a_{ig}^2}{\lambda_g}, i = 1,2,\dots, p; g = 1,2,\dots\dots, m$$

Dimana λ_g adalah akar cirri ke-j dan s^{*2} adalah galat dibagi jumlah galat terkoreksi, dirumuskan sebagai :

$$s^{*2} = \frac{s^2}{\sum (y - \bar{y})^2}$$

Pengujian signifikansi terhadap koefisien regresi secara parsial untuk mengetahui pengaruh dari setiap peubah bebas digunakan uji t-student, yaitu :

$$t(W_1) = \frac{W_1}{\sqrt{Var(W_1)}}$$

Persamaan regresi dalam bentuk peubah asal X akhirnya diperoleh sebagai berikut :

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots\dots + \alpha_p X_p$$

Contoh Kasus

Diberikan data amatan berkaitan dengan proses industrialisasi yang terjadi pada 15 kabupaten di suatu propinsi, dimana peubah-peubah dispesifikasikan terdiri dari empat peubah bebas sebagai indikator industrialisasi untuk mengetahui sejauh mana proses industrialisasi yang berlangsung (x) berpengaruh terhadap pendapatan perkapita (y) (Gazperz, 1992) sebagai berikut :

Item	y	x1	x2	x3	x4
1	67,5	9,75	6,5	1,61	0,65
2	68,9	10,5	10,25	2	0,75
3	70,65	11,25	11,9	2,5	0,9
4	73,6	12,6	11,75	2,7	1,15
5	71,89	11,9	11	2,25	0,95
6	84,5	15,2	13,5	3,25	1,75
7	72,34	12,25	12	2,9	1,05
8	77,65	12,9	12,6	3	1
9	80,25	14,3	13,2	3,1	1,7
10	79,87	13,25	12,9	3,05	1,25
11	86,75	15,3	14	3,25	1,8
12	65,75	8,9	9,25	1,9	0,6
13	70,2	10,6	10,5	1,95	0,5
14	89,25	17,25	15	3,5	2
15	85	16,9	14,9	3,4	1,95

Dimana :

Y = pendapatan perkapita (PDRB per kapita), diukur dalam satuan juta rupiah.

x_1 = kontribusi industri manufaktur dalam produk domestik regional bruto (PDRB), diukur dalam satuan persen (%).

x_2 = Banyaknya tenaga kerja dalam sektor industri manufaktur, diukur dalam satuan persen (%). (presentase dari total tenaga kerja daerah tersebut).

x_3 = produktivitas tenaga kerja industry manufaktur, diukur dalam satuan juta rupiah per tenaga kerja. (nilai tambah industry manufaktur per tenaga kerja).

x_4 = investasi dalam industri manufaktur per tenaga kerja, diukur dalam satuan juta rupiah per tenaga kerja (jumlah investasi dalam industry manufaktur dibagi dengan banyaknya tenaga kerja industry manufaktur).

Selanjutnya dilakukan analisis apakah keempat peubah bebas tersebut (x) memberikan pengaruh positip terhadap penambahan pendapatan daerah perkapita (y).

Adapun langkah-langkah analisis sebagai berikut :

1. Menggunakan bantuan softwere MINITAB 14, dilakukan pengamatan apakah terdapat korelasi antar variabel bebas. Berdasarkan korelasi pearson yang diperoleh sebagai berikut :

	X1	X2	X3
X2	0,909		
	0,000		
X3	0,933	0,952	
	0,000	0,000	
X4	0,969	0,864	0,911
	0,000	0,000	0,000

Cell Contents: Pearson correlation
P-Value

Terlihat korelasi antar variabel seluruhnya mendekati 1(besar), juga p-value < 0,05, dapat disimpulkan bahwa hal ini menunjukkan adanya korelasi antar masing-masing variabel bebas (x_1 dengan x_2, x_3 dan x_4 ; x_2 dengan x_3 dan x_4 ; x_3 dengan x_4).

2. Analisis menggunakan MKT dengan bantuan MINITAB 14 :

Regression Analysis: Y versus X1; X2; X3; X4

The regression equation is
Y = 41,7 + 2,35 X1 - 0,248 X2 + 2,05 X3 + 1,57 X4

Predictor	Coef	SE Coef	T	P	VIF
Constant	41,658	6,345	6,57	0,000	
X1	2,347	1,066	2,20	0,052	24,7
X2	-0,2483	0,8428	-0,29	0,774	12,2
X3	2,052	3,526	0,58	0,573	16,0
X4	1,569	4,492	0,35	0,734	18,1

S = 2,02669 R-Sq = 94,9% R-Sq(adj) = 92,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	763,97	190,99	46,50	0,000
Residual Error	10	41,07	4,11		
Total	14	805,05			

Source	DF	Seq SS
X1	1	761,41
X2	1	0,03
X3	1	2,03
X4	1	0,50

Dari analisis menggunakan MKT diatas, dapat dilihat bahwa berdasarkan Analisis of Variance, p-value < 0.05 dan R-Sq = 94,9% (tinggi) menunjukkan bahwa model nyata secara statistic. Sedangkang berdasarkan uji parsial p-value > 0.05, artinya tidak ada variabel yang nyata secara statistik. Akibatnya analisis menjadi tidak valid seperti teori yang telah dijelaskan sebelumnya. Hal ini mengindikasikan bahwa terdapat multikolinearitas diantara peubah bebas. Juga dari nilai VIF yang diperoleh, ternyata semuanya lebih dari 10. Sehingga disimpulkan terdapat multikolinearitas antar variabel bebas.

3. Analisis menggunakan komponen utama :

    a. Pembakuan data X.

Karena satuan antara peubah bebas tidak sama dan range antar variabel cukup besar, maka yang digunakan adalah matriks korelasi, sehingga langkah pertama yaitu membakukan data (x), diperoleh data hasil pembakuan (z) sebagai berikut :

Z1	Z2	Z3	Z4
-1,23053	-2,42986	-1,76149	-1,07131
-0,93346	-0,75794	-1,12579	-0,87652
-0,63639	-0,02229	-0,31079	-0,58435
-0,10166	-0,08917	0,01521	-0,09739
-0,37893	-0,42355	-0,71829	-0,48696
0,92818	0,69106	0,91171	1,07131
-0,2403	0,02229	0,34121	-0,29217
0,01716	0,2898	0,50421	-0,38957
0,5717	0,55731	0,66721	0,97391
0,1558	0,42355	0,58571	0,09739
0,96779	0,91398	0,91171	1,1687
-1,56721	-1,20379	-1,28879	-1,1687
-0,89385	-0,64648	-1,20729	-1,36348
1,74017	1,35983	1,31921	1,55826
1,60154	1,31525	1,15621	1,46087

b. Menentukan akar ciri, vektor ciri dan skor komponen utama untuk seluruh data.

Eigenanalysis of the Correlation Matrix

Eigenvalue	3,7694	0,1625	0,0442	0,0239
Proportion	0,942	0,041	0,011	0,006
Cumulative	0,942	0,983	0,994	1,000

Variable	PC1	PC2	PC3	PC4
Z1	-0,506	0,340	0,357	-0,708
Z2	-0,494	-0,639	0,507	0,301
Z3	-0,504	-0,318	-0,781	-0,187
Z4	-0,497	0,612	-0,075	0,611

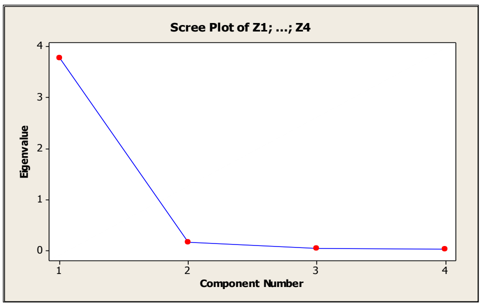
Dari hasil analisis diatas dapat diperoleh persamaan untuk masing-masing komponen utama sebagai berikut :

PC1 = – 0,506Z1 – 0,494Z2 – 0,504Z3 – 0,497Z4
PC2 = 0,340Z1 – 0,639Z2 – 0,318Z3 +0,6127Z4
PC3 = 0,357Z1 + 0,507Z2 – 0,781Z3 – 0,075Z4
PC4 = – 0,708Z1 + 0,301Z2 – 0,187Z3 + 0,611Z4

Selanjutnya dari persamaan diatas, diperoleh skor komponen utama untuk masing masing komponen utama yaitu :

SK(PC1)	SK(PC2)	SK(PC3)	SK(PC4)
3,243231	1,03882	-0,21517	-0,18534
1,849782	-0,01149	0,227459	0,107717
0,780085	-0,46092	0,048057	0,144935
0,136227	-0,04203	-0,08608	-0,01721
1,00501	0,07221	0,247484	-0,02242
-1,80299	0,23971	-0,11067	0,034936
0,083819	-0,38326	-0,31906	-0,06549
-0,21235	-0,5781	-0,21152	-0,25724
-1,3849	0,22212	-0,10749	0,233282
-0,63167	-0,34433	-0,19439	-0,03284
-1,98155	0,17033	0,00919	0,133498
2,618075	-0,06904	-0,07562	0,274175
2,057773	-0,34134	0,398283	-0,16907
-2,99162	0,25687	0,1635	-0,11733
-2,76889	0,23046	0,226011	-0,06162

Scree plot komponen utama :



Berdasarkan scree plot maupun nilai proporsi dari komponen utama, 2 komponen utama PC1 dan PC2 sudah dapat menjelaskan keragaman sebesar 98.3%, maka untuk selanjutnya PC1 dan PC2 ini sudah layak dipakai sebagai variabel baru yang saling bebas untuk analisa regresi. PC1 dan PC2 masing-masing merupakan kombinasi linear dari empat peubah asal yang telah dibakukan (z). Selanjutnya menggunakan MKT diperoleh persamaan regresi komponen utama sebagai berikut :

The regression equation is
Y = 76,3 - 3,75 SK(PC1) + 2,46 SK(PC2)

Predictor	Coef	SE Coef	T	P	VIF
Constant	76,2733	0,5115	149,12	0,000	
SK(PC1)	-3,7530	0,2726	-13,77	0,000	1,0
SK(PC2)	2,463	1,314	1,87	0,085	1,0

S = 1,98101 R-Sq = 94,2% R-Sq(adj) = 93,2%

Dari analisi diatas dapat dilihat, bahwa VIF menunjukkan tidak ada korelasi antar variabel komponen utama. Selanjutnya dilakukan uji parsial untuk melihat pengaruh dari masing-masing variabel,berdasarkan p-value yang diperoleh menunjukkan bahwa hanya PC1 yang memiliki pengaruh signifikan terhadap model, sedangkan PC2 tidak. Sehingga model bisa diperingkas dengan menghilangkan variabel PC2.

The regression equation is
Y = 76,3 - 3,75 SK(PC1)

Predictor	Coef	SE Coef	T	P
Constant	76,2733	0,5588	136,50	0,000
SK(PC1)	-3,7532	0,2978	-12,60	0,000

S = 2,16418 R-Sq = 92,4% R-Sq(adj) = 91,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	744,16	744,16	158,88	0,000
Residual Error	13	60,89	4,68		
Total	14	805,05			

SK(PC1) merupakan fungsi dari PC1, bila disubtitusikan dengan PC1 diperoleh :

$$\hat{y} = 76{,}3 - 3{,}76(-0{,}506Z1 - 0{,}494Z2 - 0{,}504Z3 - 0{,}497Z4)$$

$$\hat{y} = 76{,}3 + 1{,}517Z1 + 1{,}847Z2 + 1{,}883Z3 + 1{,}857Z4$$

Selanjutnya dilakukan uji parsial terhadap masing-masing variabel baku, untuk melihat terdapat pengaruh atau tidak darimasing-masing variabel baku.

Ditentukan terlebih dahulu $s*^2$ menggunakan rumus :

$$s^{*2} = \frac{s^2}{\sum (y-\bar{y})^2} = \frac{4.68}{805.05} = 0.0058$$

Selanjutnya diperoleh ragam koefisien regresi utama

$$Var(PC_j) = s^{*2} \sum_{g=1}^{m} \frac{a_{ij}}{\lambda_j}, i = 1,2......, p; j = 1,2,......m.$$

Karena komponen utama yang terlibat dalam regresi komponen utama hanya satu komponen utama, jadi m=1. Dengan demikian diperoleh :

$$Var(PC_1) = s^{*2} \frac{a_{i1}}{\lambda_1}$$

Dimana i=1,2,3,4 dan a_{i1} adalah koefisien komponen utama pertama (vektor ciri pertama), λ_1 adalah akar ciri pertama, sehingga dapat ditentukan ragam (variance) dari koefisien regresi $\gamma_i, i = 1,2,3,4$.

$$Var(\gamma_1) = (0.0058)\frac{(-0{,}506)^2}{3{,}7694} = 0{,}000394$$

$$Var(\gamma_2) = (0.0058)\frac{(-0{,}494)^2}{3{,}7694} = 0{,}000375$$

$$Var(\gamma_3) = (0.0058)\frac{(-0{,}504)^2}{3{,}7694} = 0{,}000391$$

$$Var(\gamma_4) = (0.0058)\frac{(-0{,}497)^2}{3{,}7694} = 0{,}000380$$

Sehingga diperoleh galat baku dari koefisien regresi baku adalah :

$$s(\gamma_1) = \sqrt{Var(\gamma_1)} = 0{,}0198$$

$$s(\gamma_2) = \sqrt{Var(\gamma_2)} = 0{,}0194$$

$$s(\gamma_3) = \sqrt{Var(\gamma_3)} = 0{,}0198$$

$$s(\gamma_4) = \sqrt{Var(\gamma_4)} = 0{,}0195$$

Uji signifikansi koefisien regresi baku adalah :

$$t(\gamma_1) = \frac{\gamma_1}{s(\gamma_1)} = \frac{1{,}517}{0{,}0198} = 76{,}6162$$

$$t(\gamma_2) = \frac{\gamma_2}{s(\gamma_2)} = \frac{1{,}847}{0{,}0194} = 95{,}2062$$

$$t(\gamma_3) = \frac{\gamma_3}{s(\gamma_3)} = \frac{1{,}883}{0{,}0198} = 95{,}1010$$

$$t(\gamma_4) = \frac{\gamma_4}{s(\gamma_4)} = \frac{1{,}875}{0{,}0195} = 95{,}2308$$

Dari persamaan regresi baku berdasarkan t hitung yang diperoleh, tampak bahwa keempat peubah bebas nyata secara statistik. Dapat disimpulkan bahwa ukuran industrialisasi memiliki peranan yang relative sama besarnya terhadap pendapatan per kapita (y).

Selanjutnya dari regresi baku, dikembalikan ke kondisi semula yaitu :

$$\hat{y} = 76{,}3 + 1{,}517\left(\frac{x_1 - 12{,}86}{2{,}52}\right) + 1{,}847\left(\frac{x_2 - 11{,}95}{2{,}24}\right) + 1{,}883\left(\frac{x_3 - 2{,}69}{0{,}61}\right) + 1{,}857\left(\frac{x_4 - 1{,}20}{0{,}51}\right)$$

$$\hat{y} = 46{,}00062 + 0{,}6019x_1 + 0{,}8247x_2 + 3{,}0870x_3 + 0{,}6418x_4$$

KESIMPULAN

Berdasarkan ilustrasi contoh kasus, menunjukkan bahwa analisis menggunakan regresi komponen utama cukup efektif dalam mengatasi masalah multikolinearitas. Ini terlihat dimana nilai VIF pada regresi komponen utama bernilai satu (menunjukkan tidak terdapat korelasi antar variabel komponen utama). Selain itu, berdasarkan uji parsial terhadap masing-masing variabel z menunjukkan bahwa masing-masing variabel berpengaruh nyata terhadap y. Sedangkan jika dilihat dari standar eror penduga koefisien regresi, pada penduga koefisien regresi komponen utama bernilai lebih kecil, sehingga bisa dikatakan lebih tepat dan lebih reliabel.

DAFTAR PUSTAKA

Aunuddin, 1989. Analisis Data. Pusat Antar Universitas Ilmu Hayat, Institut Pertanian Bogor, Bogor.

Gasperz, V. 1992. Teknik Analisis dalam Penelitian Percobaan. Tarsito. Bandung.

Jollife, I. T., 1986. Principal Componen Analysis. Springer-Verlag. Newyork.

Myers, R.H. & J.S.Milton. 1991. A First Course in the Theory of Linear Statistical Models. PWS-KENT Publishing Company. Bosto.